



**Hachuel, Leticia**  
**Boggio, Gabriela**  
**Wojdyla, Daniel**  
**Servy, Elsa**

*Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística*

## **INTERPRETACIÓN Y COMPARACIÓN DE MODELOS DE REGRESIÓN LOGÍSTICA PARA EL ESTUDIO DE LA DESOCUPACIÓN\***

### **1. INTRODUCCIÓN**

La Encuesta Permanente de Hogares (EPH) constituye un programa nacional de producción sistemática y permanente de indicadores sociales que lleva a cabo el Instituto Nacional de Estadísticas y Censos. Además de permitir conocer las características socio-demográficas y económicas de la población sirve de base para proporcionar entre otros resultados las tasas oficiales de desocupación. La información se releva a través de una muestra de tipo estratificado, con etapas múltiples de selección, según la cual se eligen en una etapa final hogares encuestándose a todos sus integrantes.

Hasta mayo del año 2003 la encuesta se realizaba dos veces por año con un esquema de rotación que imponía que en cada onda se eliminara de ella una submuestra, constituida por el 25% de los hogares, los que eran reemplazados por un número equivalente de hogares elegidos en forma independiente de modo que la muestra, después de cuatro ondas, quedaba sustituida en su totalidad. A partir de la fecha mencionada se realizaron cambios sustanciales en ciertos aspectos del diseño de la encuesta bajo una modalidad de relevamiento continuo, que abarcan tanto la periodicidad como el esquema de rotación del mismo.

La información suministrada por esta encuesta permite construir modelos estadísticos que ponen de manifiesto las relaciones entre la desocupación y variables demográficas y económicas. Ellos permiten calcular tasas de desocupación para diferentes estratos haciendo uso de toda la información disponible facilitando además especulaciones sobre situaciones posibles aunque no realmente observadas.

Por otro lado, resulta obvio reconocer al período 2001-2002 como de características críticas y en especial en su relación con el problema de la desocupación. Si bien no existen certezas respecto al momento de finalización de dicha crisis, es razonable pensar al año 2003 como de relativa superación.

En el período de crisis, 2001-2002, dadas las características de panel rotativo de la encuesta, la información disponible sobre desocupación es de carácter longitudinal, abarcando 4 ondas consecutivas. Este tipo de datos supone la existencia de correlación entre las observaciones realizadas sobre un mismo individuo a través del tiempo. Además, otra consecuencia del diseño muestral es que mientras las respuestas de individuos de diferen-

---

\* Este trabajo fue realizado en el marco del proyecto "La información estadística como base para el diagnóstico de la desocupación en Rosario" Proyecto de investigación y desarrollo científico- tecnológico de alto impacto socioeconómico. Año 2004. SPU\_UNR 0044.



tes hogares se pueden suponer independientes, las correspondientes a los distintos integrantes de un mismo hogar dependen unas de otras.

Para el año 2003, los datos referidos a sólo una onda permiten reconocer sólo asociaciones entre las respuestas de individuos de un mismo hogar.

En ambos períodos, la consideración de estas posibles fuentes de correlación en la modelización de la probabilidad de desocupación se realizó mediante el ajuste de modelos logísticos marginales (Hachuel et al, 2003; Hachuel et al, 2005).

Este trabajo pretende mostrar a partir de los modelos logísticos marginales ajustados, formas alternativas de presentación de resultados a los efectos de facilitar su comprensión y comparación.

A continuación se realiza una breve descripción de la metodología utilizada en los análisis previos y los resultados hallados en cada uno de los dos períodos señalados, para luego mostrar las distintas formas de presentación de los mismos.

## 2. ESTUDIOS PREVIOS

### 2.1 Metodología

Los objetivos principales en la modelización estadística con datos binarios correlacionados son, por un lado, adaptar a este contexto las herramientas de regresión convencionales para relacionar la variable respuesta con variables explicativas y por otro lado, tener en cuenta la correlación entre los sujetos.

Al igual que para datos binarios independientes, los modelos de regresión logística (Hosmer, Lemeshow, 2000) son los más populares para datos binarios correlacionados (Aerts et al., 2002).

El enfoque conocido como Ecuaciones de Estimación Generalizadas (GEE), introducido por Liang y Zeger (1986), es un método que provee una estrategia válida para cumplir con ambos objetivos.

Dicho enfoque modela una función conocida de la esperanza marginal de la variable dependiente como una función lineal de una o más variables explicativas. Los coeficientes de la regresión marginal tienen la misma interpretación que los coeficientes de regresión de un análisis transversal. Más formalmente, se asume que las mediciones repetidas proveen una respuesta multivariada:  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ij}, \dots, Y_{ik_i})'$  correspondiente al  $i$ -ésimo sujeto, donde  $i=1, \dots, n$  y en cada ocasión se observa la respuesta  $Y_i$  como así también un vector  $p \times 1$  de variables explicativas  $x_{ik_i}$ .

Se modelan separadamente la regresión y la correlación intra sujeto suponiendo:

1. la esperanza marginal de  $Y_{ij}$ ,  $E(Y_{ij}) = \mu_{ij}$ , se relaciona con  $x_{ij}$  por  $g(\mu_{ij}) = x_{ij}'\beta$ , donde  $g$  es una función de enlace conocida, tal como la función logit para respuestas binarias;

2. la variancia marginal es una función de la media marginal, esto es,  $\text{var}(Y_{ij}) = v(\mu_{ij})\phi$ , donde  $v$  es una función conocida y  $\phi$  es el parámetro de sobredispersión que toma en cuenta la variación de  $Y_{ij}$  no explicada por  $v(\mu_{ij})$ ;

3. la covariancia entre  $Y_{ij}$  e  $Y_{ij'}$  es una función de las medias marginales y de parámetros

adicionales  $\alpha$ ; esto es:  $\text{cov}(Y_{ij}, Y_{ij'}) = d(\mu_{ij}, \mu_{ij'}, \alpha)$  donde  $d$  es una función conocida.

En particular, para respuesta binaria, donde  $\mu_{ij} = \pi_{ij} = P(Y_{ij} = 1)$ , el modelo marginal más ampliamente usado es el que toma como función de enlace la transformación logit, esto es:

$$\text{logit } \pi_{ij} = \ln \frac{\pi_{ij}}{1 - \pi_{ij}} = x'_{ij} \beta. \quad (1)$$

A partir del ajuste del modelo es posible estimar:

. probabilidades de respuesta positiva para valores específicos de las covariables:

$$\hat{\pi} = \frac{1}{(e^{-c} + 1)}, \quad (2)$$

siendo  $c$  el "logit" estimado para dichos valores específicos de las covariables,

. el odds que se interpreta como chance o riesgo:

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = e^c \text{ y} \quad (3)$$

. la razón de "odds" entre la respuesta y una de las covariables, suponiendo fijas las restantes:

$$\hat{\theta} = \frac{\hat{\pi}_1 / (1 - \hat{\pi}_1)}{\hat{\pi}_2 / (1 - \hat{\pi}_2)} = e^{c_1 - c_2}, \quad (4)$$

siendo  $\hat{\pi}_1$  y  $\hat{\pi}_2$  las probabilidades estimadas para los mismos valores específicos de las covariables que no intervienen en la razón de "odds".

Con respecto a la asociación entre las respuestas de un mismo individuo a través del tiempo, la mejor forma de expresarla, cuando la variable respuesta es binaria, es a través de la razón de odds entre las respuestas de ese individuo en dos momentos distintos,  $j$  y  $j'$ :

$$\Psi_{ijj'} = \frac{P(Y_{ij} = Y_{ij'} = 1) P(Y_{ij} = Y_{ij'} = 0)}{P(Y_{ij} = 1; Y_{ij'} = 0) P(Y_{ij} = 0; Y_{ij'} = 1)} \quad (5)$$

Esta alternativa fue sugerida por Carey et al. (1993) bajo la denominación de "Alternating Logistic Regressions" (ALR). Este método ajusta, en forma alternada, el modelo (1) y un modelo para las razones de "odds" entre pares de respuestas:

$$\log \Psi_{ijj'} = z'_{ijj'} \alpha, \quad (6)$$

donde  $z_{ijj'}$  es un vector de dimensión  $q \times 1$  de covariables que especifican la forma de la asociación entre  $Y_{ij}$  e  $Y_{ij'}$ . El logaritmo de la razón de odds de dos observaciones cualesquiera de diferentes grupos se supone igual a cero, lo que implica independencia entre ellos.

El algoritmo admite estructuras complejas para los logaritmos de las razones de odds, incluyendo asociaciones diferentes para subgrupos dentro de un grupo de mediciones. Este enfoque es útil cuando la estructura de asociación es en sí también misma de interés.

En síntesis, el enfoque de Carey et al. (1993) caracteriza la dependencia entre respuestas en términos de razones de odds y usa las GEE para estimar los parámetros de re-



gresión alternando con regresiones logísticas de cada respuesta sobre otras en el mismo grupo para actualizar las estimaciones de los parámetros razones de "odds".

## 2.2 Resultados

### *Período 2001-2002*

Para el estudio de la tasa de desocupación en el Gran Rosario a través de modelos marginales se utiliza información provista por la EPH para las ondas de los años 2001 y 2002 referida a la población económicamente activa (PEA), es decir personas que trabajan o buscan trabajo con edades comprendidas entre 15 y 65 años de edad

Para las cuatro ondas del período 2001-2002, se pueden obtener paneles independientes de diferentes duraciones, medidos en distintas ocasiones, como lo indica el esquema siguiente:

Tabla 1: Esquema de rotaciones para el período mayo 01/octubre 02

Rotación	Onda			
	Mayo 01	Octubre 01	Mayo 02	Octubre 02
1	X	X	X	X
2	X	X	X	-
3	-	X	X	X
4	X	X	-	-
5	-	-	X	X
6	X	-	-	-
7	-	-	-	X

De acuerdo al esquema, los miembros de cada hogar en el período bajo estudio pueden haber sido entrevistado una, dos, tres o cuatro veces. En el análisis estadístico se tiene en cuenta el doble agrupamiento de la información: por un lado los individuos están agrupados en familias y por otro, cada individuo proporciona información repetidamente en a lo sumo las 4 ondas consideradas.

En base a la información descripta se construye un modelo capaz de explicar la tasa de desocupación a partir de un conjunto de variables explicativas relacionadas con características personales y laborales. La elección de dichos factores se basa en argumentos económicos, los cuales han sido descriptos en trabajos previos (Servy et al., 2000). Ellos se traducen en las siguientes variables, a saber:

Sexo: femenino, masculino.

Edad: variable continua medida en años.

Escolaridad: primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, nivel superior o universitario incompleto y nivel superior o universitario completo.

Nivel de ingreso: bajo (1º, 2º y 3º decil de ingreso), medio (4º a 7º decil) y alto (8º a 10º decil).

Rama de actividad: construcción, manufacturera, servicios comerciales y de transporte, intermediación financiera, administración pública o defensa e instrucción pública o servicios de salud, otros servicios.

Tamaño de la empresa: unipersonal, de 2 a 5 personas, de 6 a 25 personas, 26 a 100



personas y más de 100 personas.

Además, en este enfoque longitudinal, la modelización adoptada requiere la definición de una covariable que identifique el momento de obtención de la información –onda- dentro del período bajo estudio. La misma se considera medida a escala nominal, con categoría de referencia mayo 01.

Con el objeto de tener en cuenta en el proceso de modelización tanto la asociación de las respuestas de un mismo individuo a través del tiempo como también la correlación inducida por estar los individuos agrupados en familias, se especifica la siguiente estructura de asociaciones entre las observaciones. Establece una razón de odds constante entre las respuestas de pares de individuos de una misma familia,  $\alpha_0$  (asociación entre individuos) y un valor constante para las razones de odds entre las respuestas de un mismo individuo en todos los pares de ondas,  $\alpha_0 + \alpha_1$  (asociación intra-individuos), donde  $\alpha_1$  es el incremento en la asociación debido a la observación de respuestas repetidas en el tiempo del mismo individuo.

El modelo marginal ajustado, con dicha estructura de asociación, incluye las covariables categóricas antes definidas y edad como efectos principales. Respecto de esta última covariable, medida en escala continua, se modela a través de una componente lineal y otra cuadrática a fin de captar los diferentes cambios en la probabilidad de desocupación según sea la edad.

Los resultados hallados señalan como efectos significativos en forma global los referidos a escolaridad, nivel de ingreso, rama de actividad y onda; no así sexo ni tamaño de la empresa (Tabla A1).

En relación a la estructura de asociación utilizada en el enfoque ALR, la falta de significancia de la asociación familiar sugiere que los factores considerados en el modelo para la probabilidad de desocupación alcanzan para explicar la asociación entre individuos de una misma familia. Para ahondar en esta línea se ajustó un modelo para la probabilidad de desocupación sin covariables y el mismo proporciona asociaciones entre individuos (intra-hogar) e intra-individuos significativas. Más aún si se ajusta un modelo con todas las covariables salvo nivel de ingreso familiar, también conduce a valores significativos en ambas asociaciones. Este hecho podría explicarse por ser esta covariable una característica medida a nivel hogar e imputada con el mismo valor a cada integrante del mismo hogar. En cambio la asociación intra-sujeto no alcanza a mitigarse por la inclusión de variables en el modelo marginal. El signo positivo de dicha asociación sugiere persistencia de los individuos en el estado ocupacional a través del tiempo.

### ***Onda mayo 2003***

El ajuste del modelo para este corte transversal se realiza también mediante la metodología ALR descripta en la sección 2.1. Para ello es necesario determinar una estructura de asociación factible entre las observaciones. Se elige especificar una razón de odds constante entre las respuestas de pares de individuos de un mismo hogar. El modelo marginal ajustado con dicha estructura de asociación presenta los efectos principales asociados a sexo, edad, escolaridad, rama de actividad y tamaño de la empresa.

En la Tabla A2 se presentan las estimaciones de los parámetros de regresión y de asociación. Con respecto a los parámetros de regresión, resultan significativos en forma global los relacionados con la escolaridad y rama de actividad de los individuos. Si bien esto no ocurre con el test global para edad, las pruebas univariadas de Wald correspondientes a



sus dos componentes resultan levemente significativos. Resta mencionar que ni el sexo ni el tamaño de la empresa donde trabaja o trabajó el individuo tuvieron un efecto significativo sobre la probabilidad de desocupación.

En relación a la estructura de dependencia, la leve contundencia en la falta de significación de la asociación intra-hogar ( $p=0.05981$ ), indicaría que en este caso los factores considerados en el modelo no resultan del todo suficientes para explicar la asociación entre individuos de un mismo hogar.

### 3. ESTUDIO COMPARATIVO

En base a los resultados hallados en los análisis recién descriptos y a los efectos de la comparación se reajustan ambos modelos con las siguientes modificaciones. No se considera la covariable tamaño de la empresa por su falta de significación; se incluye, sin embargo, la covariable sexo a pesar de tampoco ser significativa por su relevancia a la hora de estimar tasas específicas, y por último se recategoriza la variable escolaridad con el objeto de disminuir el conjunto de estratos a caracterizar. Por otro lado, la leve asociación intra-hogar en el modelo correspondiente a la onda mayo 2003 motiva la decisión de ajustar para este período un modelo logístico convencional. Los resultados hallados se presentan en la tabla 2.

En ambos períodos resulta significativo el efecto de la edad y del nivel de ingreso, no así del sexo. En relación al nivel de escolaridad se observan algunas diferencias: en el período 2001-2002 sólo resulta significativo el efecto sobre la probabilidad de desocupación de un alto nivel de escolaridad (superior o universitario completo) respecto del nivel más bajo; en cambio en mayo 2003, resultan significativos los efectos asociados a todos los niveles. En relación a rama de actividad, la única diferencia se manifiesta en la pérdida de significación y cambio de signo del efecto de la rama manufacturera en comparación con la construcción para el año 2003.

Si bien el modelo marginal ajustado para las cuatro ondas (2001-2002) no presenta diferencias significativas en la comparación entre ellas, podría existir algún grado de incertidumbre en la relación entre mayo 2002 y mayo 2001 ( $p\text{-asoc}=0.083$ ), hecho que da lugar a señalar a la onda correspondiente a mayo 2002 como la más representativa del período de crisis considerado (Hachuel et al, 2003).

La interpretación de los coeficientes en un modelo de regresión logística comúnmente se realiza mediante cambios en la escala "logit" o a través de razones de "odds" condicionales. Sin embargo dichos coeficientes no tienen una interpretación directa en términos de probabilidad, que podría considerarse como la preferida para la mayoría de los usuarios.

Una forma simple de utilizar el modelo es estimar la probabilidad de desocupación para perfiles poblacionales definidos a partir de combinaciones de valores de las covariables que intervienen en el modelo.



Tabla 2: Estimación de los coeficientes de los modelos de regresión logística para el período 2002-2002 y para mayo 2003

Variable	2001- 2002		Mayo 2003	
	coeficiente	p-asoc	coeficiente	p-asoc
Sexo				
Masculino	-	-	-	-
Femenino	0.164	0.278	0.689	0.118
Edad				
Comp. lineal	-0.118	<b>&lt;0.0001</b>	-0.206	<b>0.017</b>
Comp. Cuadrática	0.001	<b>&lt;0.0001</b>	0.003	<b>0.015</b>
Escolaridad				
Primario incomp.	-	-	-	-
Prim. comp./Secund. inc.	0.034	0.879	1.429	<b>0.038</b>
Secund. comp./Sup. inc.	0.434	0.080	2.029	<b>0.009</b>
Superior o univ. comp.	0.741	<b>0.023</b>	2.467	<b>0.023</b>
Nivel de Ingreso				
Bajo	-	-	-	-
Medio	-1.337	<b>&lt;0.0001</b>	-1.614	<b>0.0001</b>
Alto	-2.750	<b>&lt;0.0001</b>	-3.704	<b>&lt;0.0001</b>
Rama de actividad				
Construcción	-	-	-	-
Manufactura	-0.636	<b>0.002</b>	1.219	0.092
Servicios comerciales	0.681	<b>0.001</b>	-1.147	<b>0.028</b>
Intermed. Financiera	-0.393	0.164	-1.271	0.123
Adm.Púb y Def+Inst Púb	-1.422	<b>&lt;0.0001</b>	-2.752	<b>&lt;0.0001</b>
Otras Activ. de Servicios	-1.072	<b>&lt;0.0001</b>	-2.323	<b>0.001</b>
Onda				
Mayo 01	-	-	-	-
Octubre 01	-0.036	0.765		
Mayo 02	0.220	0.083		
Octubre 02	-0.178	0.197		

Otra forma de presentar los resultados del modelo de regresión logística es a través de una estimación ajustada de las probabilidades asociadas a cada covariable. Este ajuste es útil especialmente en estudios observacionales donde es frecuente el imbalance de factores influyentes. En estos casos la estimación ajustada puede ser sustancialmente diferente a la no ajustada. El cálculo de probabilidades ajustadas puede ser más convincente para una audiencia no experta en las nociones de estimación de parámetros de regresión y probabilidad asociada.

A los efectos de ejemplificar estas diferentes formas de interpretación del modelo de regresión logística se presentan, en primer lugar, algunas razones de "odds" ajustadas entre estado ocupacional y rama de actividad.

Así para mayo 2002, la chance de desocupación para las personas que trabajan o han trabajado en la rama manufacturera es ( $\exp(-0.636)=0.64$ ) casi un 40 % menor que para los



pertenecientes a la rama de la construcción. En cambio, en mayo 2003, esta chance aumenta ( $\exp(1.219)=6.15$ ) seis veces con respecto a la de la construcción.

Se estiman también a modo de ejemplo probabilidades de desocupación correspondientes a diferentes ramas de actividad para un perfil poblacional particular en las ondas mayo 2002 (en base al primer modelo) y mayo 2003. Este perfil corresponde a los varones de 30 años con el nivel más bajo de escolaridad y bajos ingresos (Tabla 3).

Tabla 3: Probabilidades de desocupación para un perfil poblacional

Rama de actividad	Mayo 2002	Mayo 2003
Construcción	0.394	0.203
Manufactura	0.256	0.463
Adm.Púb y Def+Inst Púb	0.136	0.020

Es notoria la disminución de la probabilidad de desocupación estimada para el perfil poblacional estudiado en la rama de la construcción y en administración e instrucción pública y defensa. En cambio, en correspondencia con el cambio de signo ya mencionado en el coeficiente estimado para la rama manufacturera, se aprecia un aumento en la probabilidad de desocupación para dicha rama en la onda correspondiente a mayo 2003.

Una forma de independizar este comportamiento de un perfil poblacional particular se logra a través del cálculo de las respectivas probabilidades ajustadas. Ellas son el promedio ponderado de las probabilidades de desocupación para cada nivel de una covariable en particular calculada a través de todos los estratos definidos por las combinaciones de niveles de las restantes covariables. La ponderación está dada por el número de individuos en cada estrato.

La tabla 4 muestra que estas probabilidades ajustadas para diferentes ramas de actividad presentan el mismo comportamiento descripto para la tabla anterior.

Tabla 4: Probabilidades de desocupación ajustadas

Rama de actividad	Mayo 2002	Mayo 2003
Construcción	0.64	0.44
Manufactura	0.53	0.65
Adm.Púb y Def+Inst Púb	0.33	0.08

Cabe señalar que dichas probabilidades se dicen ajustadas estadísticamente en el sentido que estiman lo que se podría esperar observar si los sujetos sólo difirieran en la característica particular examinada y tuviesen igual distribución en las restantes covariables. Estas probabilidades ajustadas se pueden interpretar como probabilidades de desocupación de un individuo extraído al azar de una población de referencia dada por el espacio de las covariables.





#### 4. CONSIDERACIONES FINALES

La puesta a prueba de metodologías específicas para datos correlacionados resultó apropiada para el análisis de la desocupación en Rosario utilizando información de la EPH. Los modelos utilizados permitieron el uso eficiente y apropiado de los datos, en la medida en que la inferencia sobre las covariables se ajusta por la correlación longitudinal y/o intra-hogar, de una manera que resulta consistente con la forma en que fueron obtenidos los datos.

Una vez elegido un modelo apropiado para las características particulares de los datos, el paso siguiente impone una interpretación adecuada de los datos. Si bien los investigadores con experiencia en el análisis de regresión logística pueden sentirse cómodos con el uso de razones de odds o sus logaritmos, la interpretación en términos de probabilidades resulta más directa y fácil de visualizar.

Cabe señalar que los valores de las probabilidades ajustadas de desocupación para diferentes niveles de una covariable, por ejemplo rama de actividad, pueden representar diferencias sustanciales con las obtenidas a partir de los datos brutos ya que estas últimas ocultan el efecto sobre la desocupación de otras covariables.

Si bien en esta oportunidad se presentan a modo ilustrativo formas alternativas de interpretación de algunos resultados y, dada la relevancia del tema, está en elaboración una publicación sobre los procedimientos y alcances de la elaboración estadística del dato bruto.

#### REFERENCIAS

- Aerts, M.; Geys, H.; Molenberghs, G.; Ryan, L. 2002. *Topics in Modelling of Clustered Data*. Chapman & Hall/CRC.
- Agresti, A. 2002. *Categorical Data Analysis*. John Wiley & Sons.
- Carey, V.; Zeger, S. DIGGLE, P. 1993. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80(3): 517-526.
- Diggle, P. J.; Heagerty, P. ; Liang K-Y.; Zeger S. L. 2002. *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.
- Hachuel, L.; Boggio, G.; Méndez, F. 2005. Modelos de regresión logística para el estudio de la desocupación: enfoques promedio poblacional y grupo específico. Décimas Jornadas Investigaciones en la Facultad de Ciencias Económicas y Estadística (en prensa).
- Hachuel, L.; Boggio, G.; Méndez, F. 2003. Modelos Marginales para el Estudio de la Desocupación en Rosario 2001-2002. [CD-ROM] XXXI Coloquio Argentino de Estadística. Sociedad Argentina de Estadística.
- Hardin, W. H.; Hilbe, J. M. 2003. *Generalized Estimating Equations*. Chapman & Hall.
- Hosmer, D. W. & Lemeshow, S. 2000. *Applied Logistic Regression*, 2nd ed. John Wiley & Sons. New York.
- Liang, K.; Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.
- SAS Institute, Inc. 1999. SAS/STAT user's guide, version 8. Cary, NC :SAS Institute, Inc.
- Servy, E.; Hachuel, L.; Boggio, G.; Cuesta, C; Leone, G. 2000. Modelos estadísticos para el estudio de la desocupación. Parte I: Modelos para cortes transversales. *Cuadernos del*



IITAE N° 5. Escuela de Estadística. Facultad de Ciencias Económicas y Estadística.  
Universidad Nacional de Rosario.

Zeger, S. L.; Liang, K. 1992. An overview of methods for the analysis of longitudinal data.  
*Statistics in Medicine*, 11: 1825-1839.



## ANEXO

Tabla A1: Estimaciones de los parámetros del modelo logístico marginal para el período 2001-2002

Parámetro	Estimación	Prob - asociada
Constante	1.553	0.008
Sexo		
Masculino	-	-
Femenino	0.122	0.456
Edad		<b>0.000</b>
Edad	-0.105	<b>0.000</b>
Edad cuadrado	0.001	<b>0.000</b>
Escolaridad		<b>0.001</b>
Primario incomp.	-	-
Primario comp.	-0.108	0.657
Secundario incomp.	0.264	0.333
Secundario comp.	0.237	0.382
Superior o univ. incomp.	0.979	<b>0.002</b>
Superior o univ. comp.	0.705	<b>0.039</b>
Nivel de Ingreso		<b>0.000</b>
Bajo	-	-
Medio	-1.452	<b>0.000</b>
Alto	-2.858	<b>0.000</b>
Rama de actividad		<b>0.000</b>
Construcción	-	-
Manufactura	-0.448	<b>0.046</b>
Servicios comerciales	-0.773	<b>0.000</b>
Intermed. Financiera	-0.399	0.192
Adm.Púb y Def+Inst Púb	-1.318	<b>0.000</b>
Otras Activ. de Servicios	-1.012	<b>0.000</b>
Tamaño de la empresa		0.542
Unipersonal	-	-
2 a 5 personas	0.167	0.311
6 a 25 personas	0.161	0.419
26 a 100 personas	0.202	0.397
101 o más personas	-0.227	0.491
Onda		<b>0.031</b>
Mayo 01	-	-
Octubre 01	-0.001	0.995
Mayo 02	0.240	0.092
Octubre 02	-0.237	0.147
Log asoci. entre individuos de una familia	0.067	0.765
Log asoci. intra-individuos	2.002	<b>0.000</b>



Tabla A 2: Estimaciones de los parámetros del modelo logístico para mayo 2003

<b>Parámetro</b>	<b>Estimación</b>	<b>Prob-asociada</b>
Constante	2.7449	<b>0.0462</b>
Sexo		
Masculino	0	
Femenino	0.4715	0.1521
Edad		0.1249
Edad	-0.1272	<b>0.0415</b>
Edad al cuadrado	0.0015	<b>0.0439</b>
Escolaridad		<b>0.0030</b>
Primaria incompleta	0	-
Primaria completa	0.3068	0.1537
Secundaria incompleta	-0.0937	0.8611
Secundaria completa	-0.4686	0.3935
Superior o univ. incompleta	1.4195	<b>0.0186</b>
Superior o univ. completa	-0.4318	0.5937
Rama de actividad		0.0000
Manufactura	0	-
Construcción	-1.4156	<b>0.0173</b>
Servicios comerciales	-2.2626	<b>0.0004</b>
Intermediación financiera	-3.2905	<b>0.0006</b>
Adm.Pública yDef. elns. Pub.	-3.8855	<b>0.0000</b>
Otras activ. de servicio	-2.9819	<b>0.0000</b>
Tamaño de la empresa		0.1605
Unipersonal	0	-
2-5 personas	0.3309	0.3639
6-25 personas	-0.4969	0.3351
26-100 personas	0.8187	0.1410
101 o más personas	-1.0559	0.2183
Log asoc.intra-hogar	1.2128	0.0591